

# A Heteroscedastic Accelerated Failure Time Model for Survival Data

YIFAN WANG<sup>\*</sup>, TIAN YOU<sup>\*</sup>, MARTIN LYSY<sup>†</sup>

University of Waterloo

August 14, 2015

## Abstract

While the Cox Proportional Hazard model is a fundamental tool in survival analysis, its semi-parametric nature precludes the estimation of upper survival quantiles in the presence of heavy censoring. In contrast, fully parametric models do not suffer from this issue – at the expense of additional modeling assumptions. In this article, we extend a popular family of parametric models which make the Accelerated Failure Time (AFT) assumption to account for heteroscedasticity in the log-survival times. This adds substantial modeling flexibility, and we show how to easily and rapidly compute maximum likelihood estimators for the proposed model in the presence of censoring. In an application to the analysis of a colon cancer study, we found that heteroscedastic modeling greatly diminished the significance of outliers, while even slightly decreasing the average size of prediction intervals.

*Keywords:* Accelerated Failure Time assumption, Heteroscedastic modeling, Right-censored lifetimes, Expectation-Conditional Maximization

## Introduction

When modeling the impact on failure times  $T$  of potential predictors  $\mathbf{X} = (X_1, \dots, X_D)$ , statisticians have a number of tools at their disposal. Perhaps the most famous is the Cox Proportional Hazards (CPH) model (Cox, 1972). The CPH model is highly flexible, straightforward to fit, and accommodates censored survivals. However, the semi-parametric CPH approach typically used in practice cannot estimate the conditional survival function  $S(t|\mathbf{x}) = P(T > t | \mathbf{X} =$

---

<sup>\*</sup>Equally contributing authors.

<sup>†</sup>Email: mlysy@uwaterloo.ca.

$\mathbf{x}$ ) for  $t$  greater than the largest observed survival time (Moeschberger and Klein, 1985). This becomes an important concern when the censoring rate is high (e.g., Sy and Taylor, 2000).

In contrast, fully parametric models do not suffer from this issue. A popular family of parametric models make the Accelerated Failure Time (AFT) assumption, namely that the conditional distribution of the survival times is

$$\log(T) = \mu(\mathbf{X}) + \varepsilon,$$

where  $\varepsilon \sim f_0(t)$  is a random variable which does not depend on  $\mathbf{X}$  (Wei, 1992; Kalbfleisch and Prentice, 2002). AFT models have an appealing interpretation: the relation between the conditional survival function  $S(t|\mathbf{x})$  of  $T$  and the “baseline” survival function  $S_0(t)$  of  $\varepsilon$  is simply

$$S(t|\mathbf{x}) = S_0(\lambda(\mathbf{x}) \cdot t), \quad \text{where } \lambda(\mathbf{x}) = e^{-\mu(\mathbf{x})}.$$

However, as with any parametric model, incorrect specification of  $\mu(\mathbf{x})$  and  $f_0(t)$  can adversely affect inferential results.

The purpose of this article is to relax the AFT model’s homoscedasticity assumption on the log-survivals. Much work has been done on this in the context of random individual-level effects, referred to in this literature as “frailty modeling” (Hougaard, 1991; Keiding et al., 1997; Pan, 2001; Zhang and Peng, 2009). We adopt instead a conditionally heteroscedastic approach by considering a model of the form

$$\log(T) = \mu(\mathbf{X}) + \sigma(\mathbf{X}) \cdot \varepsilon. \quad (1)$$

Estimation for location-scale type regression models such as (1) has been extensively studied; see Müller and Stadtmüller (1987); Cai and Wang (2008) for non-parametric and Hsieh (1996); Zeng and Lin (2007); Zhang and Davidian (2008); Su et al. (2012) for semi-parametric approaches. Indeed, model (1) can be viewed as a quantile regression model (Koenker and Bassett, 1978; Koenker, 2005). For specific  $\mu(\mathbf{x})$  and  $\sigma(\mathbf{x})$ , quantile regression estimates have been developed to account for right-censoring (Powell, 1986; Portnoy, 2003) and applied to survival data (Peng and Huang, 2008). One drawback of many quantile regression models which do not specify the distribution of  $\varepsilon$  is the difficulty of constructing confidence intervals for the model parameters and quantile estimates: see for instance Koenker (1994); Angelis et al. (1993); and Kocherginsky et al. (2005) for a review of several existing methods.

Fully parametric approaches to (1) have been studied by e.g., Boscardin and Gelman (1996); Smyth (2002). Following these authors, we consider the model formulation

$$\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}, \quad \sigma^2(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\gamma}), \quad \varepsilon \sim \mathcal{N}(0, 1). \quad (2)$$

While the adequacy of a specific failure time model undoubtedly varies from one dataset to another, here we shall advocate that the Heteroscedastic Accelerated Failure Time (HAFT) model described by (2) is an attractive addition to the survival modeling toolkit for a number of reasons:

1. *Interpretability.* As with the homoscedastic AFT model, the conditional survival function of the heteroscedastic HAFT model can be obtained by a simple transformation of the baseline survival function of  $\varepsilon$ :

$$S(t|\mathbf{x}) = S_0(\lambda(\mathbf{x}) \cdot t^{\alpha(\mathbf{x})}), \quad \text{where} \quad \begin{aligned} l(\mathbf{x}) &= e^{-\mu(\mathbf{x})/\sigma(\mathbf{x})}, \\ \alpha(\mathbf{x}) &= 1/\sigma(\mathbf{x}). \end{aligned}$$

For the HAFT model (2) we are proposing, it is easy to evaluate  $S(t|\mathbf{x})$  for any combination of  $t$  and  $\mathbf{x}$  using the quantile function of a standard Normal distribution.

2. *Tractability.* The HAFT model enjoys a simple algorithm for computing maximum likelihood estimators of  $\beta$  and  $\gamma$  (e.g., Smyth, 1989; Verbyla, 1993) – full details and an implementation using standard statistical software are provided in Section 1. Moreover, confidence intervals for the model parameters and quantile estimates can readily be constructed from the Hessian matrix of the log-likelihood.
3. *Censoring.* The HAFT model (2) admits a simple Expectation-Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993) to estimate  $\beta$  and  $\gamma$  in the presence of right-censored failure times (described in Section 1.2).
4. *Flexibility.* As a generalization of the homoscedastic case, the HAFT model adds considerable flexibility to the modeling of failure times. We illustrate this with data from the well-known colon cancer clinical trial of Laurie et al. (1989). The HAFT model was found to have far fewer outliers than its homoscedastic counterpart, while actually *decreasing* the average size of prediction intervals.

Elaborating on these points, the remainder of this article is organized as follows. Parametric estimation for the HAFT model in the presence of censoring is detailed in Section 1. A comparison of its performance on the colon cancer data relative to the homoscedastic AFT model is presented in Section 2. We conclude with a discussion of further work in Section 3.

## 1 Parameter Estimation for the HAFT Model

Let  $R_i = \log(T_i)$  and  $\mathbf{X}_i = (X_{i1}, \dots, X_{iD})$  denote the log-survival time and predictors for subject  $i$ . For ease of exposition, we decompose the covariates of the HAFT model into their mean and variance effects:

$$R_i | \mathbf{X}_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\mathbf{W}_i' \beta, \exp(\mathbf{Z}_i' \gamma)), \quad (3)$$

where  $\mathbf{W}_i = (W_{i1}, \dots, W_{ip}) = \mathbf{f}(\mathbf{X}_i)$  and  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iq}) = \mathbf{g}(\mathbf{X}_i)$ . The model parameters are  $\beta = (\beta_1, \dots, \beta_p)$  and  $\gamma = (\gamma_1, \dots, \gamma_q)$ , and the log-

likelihood function is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{R}, \mathbf{X}) = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{(R_i - \mathbf{W}_i' \boldsymbol{\beta})^2}{\exp(\mathbf{Z}_i' \boldsymbol{\gamma})} + \mathbf{Z}_i' \boldsymbol{\gamma} \right],$$

where  $\mathbf{R} = (R_1, \dots, R_n)$ .

### 1.1 Maximum Likelihood Estimation Without Censoring

We first present a method of calculating the MLE of  $(\boldsymbol{\beta}, \boldsymbol{\gamma})$  for complete (uncensored) cases. For fixed  $\boldsymbol{\gamma}$ , the conditional log-likelihood for the mean parameters is

$$\ell(\boldsymbol{\beta} | \boldsymbol{\gamma}, \mathbf{R}, \mathbf{X}) = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{(R_i - \mathbf{W}_i' \boldsymbol{\beta})^2}{\sigma_i^2} \right], \quad \text{where } \sigma_i^2 = \exp(\mathbf{Z}_i' \boldsymbol{\gamma}).$$

This is the log-likelihood function of a Normal linear model with known variances  $\sigma_i^2$ . With  $\mathbf{W}_{n \times p} = [\mathbf{W}_1 | \dots | \mathbf{W}_n]'$ , it is maximized at

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}' \boldsymbol{\Omega} \mathbf{W})^{-1} \mathbf{W}' \boldsymbol{\Omega} \mathbf{R}, \quad \text{where } \boldsymbol{\Omega}^{-1} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2). \quad (4)$$

For fixed  $\boldsymbol{\beta}$ , the conditional log-likelihood of the variance parameters is

$$\ell(\boldsymbol{\gamma} | \boldsymbol{\beta}, \mathbf{R}, \mathbf{X}) = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{U_i}{\exp(\mathbf{Z}_i' \boldsymbol{\gamma})} + \mathbf{Z}_i' \boldsymbol{\gamma} \right], \quad \text{where } U_i = (R_i - \mathbf{W}_i' \boldsymbol{\beta})^2. \quad (5)$$

This has long been recognized as the log-likelihood of a Generalized Linear Model (GLM) for a Gamma distribution with logarithmic link function (e.g., Nelder and Pregibon, 1987; Smyth, 1989). The latter provides a Fisher scoring algorithm which iteratively updates  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  and converges to the MLE (Smyth, 1989). While further accelerations are possible (e.g., Smyth, 2002), the maximization of GLM likelihoods at present can be easily accomplished with tools from standard regression software. For example, with  $\mathbf{U}_{n \times 1} = (U_1, \dots, U_n)$  and the matrix  $\mathbf{Z}_{n \times q} = [\mathbf{Z}_1 | \dots | \mathbf{Z}_n]$ , the maximum  $\hat{\boldsymbol{\gamma}}$  of (5) can be computed in R with the command

$$\text{glm}(\mathbf{U} \sim \mathbf{Z} - 1, \text{family} = \text{Gamma}("log")). \quad (6)$$

We found that alternating between updates (4) and (6) converged very quickly to the MLE of  $(\boldsymbol{\beta}, \boldsymbol{\gamma})$  in the data analysis of Section 2.

### 1.2 An ECM Algorithm for Censored Observations

A common feature of lifetime survival data is the censoring of observations. Instead of observing the actual (log) failure time  $R_i$ , we observe  $Y_i$ , where  $Y_i = \min(R_i, C_i)$  and  $C_i$  is the censoring time. We also observe  $\delta_i = \mathbf{1}\{R_i < C_i\}$ ,

an indicator variable for whether the survival time of subject  $i$  is censored or not ( $\delta_i = 1$  means uncensored). Assuming that  $R$  and  $C$  are conditionally independent given the covariates, here we describe an Expectation-Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993) which extends the well-known Expectation Maximization algorithm (e.g., Aitkin, 1981) for the censored homoscedastic linear model to our heteroscedastic setting.

Let  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ , and  $(\boldsymbol{\beta}^{(t)}, \boldsymbol{\gamma}^{(t)})$  denote the parameter values at iteration  $t$ .

- *E-step*: We have

$$\begin{aligned} Q_t(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= E[\ell(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{R}, \mathbf{X}) \mid \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X}, \boldsymbol{\beta}^{(t)}, \boldsymbol{\gamma}^{(t)}] \\ &= E \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(R_i - \mathbf{W}_i' \boldsymbol{\beta})^2}{\exp(\mathbf{Z}_i' \boldsymbol{\gamma})} - \frac{1}{2} \sum_{i=1}^n \mathbf{Z}_i' \boldsymbol{\gamma} \mid \mathbf{Y}, \boldsymbol{\delta}, \mathbf{X}, \boldsymbol{\beta}^{(t)}, \boldsymbol{\gamma}^{(t)} \right] \\ &= -\frac{1}{2} \sum_{i=1}^n \frac{\tilde{S}_i - 2\tilde{R}_i \mathbf{W}_i' \boldsymbol{\beta} + (\mathbf{W}_i' \boldsymbol{\beta})^2}{\exp(\mathbf{Z}_i' \boldsymbol{\gamma})} - \frac{1}{2} \sum_{i=1}^n \mathbf{Z}_i' \boldsymbol{\gamma}, \end{aligned}$$

where

$$\tilde{R}_i = \begin{cases} Y_i & \delta_i = 1 \\ \sigma_i^{(t)} f(\tilde{Y}_i) + \mu_i^{(t)} & \delta_i = 0, \end{cases} \quad \tilde{S}_i = \begin{cases} Y_i^2 & \delta_i = 1 \\ (\sigma_i^{(t)})^2 g(\tilde{Y}_i) + 2\mu_i^{(t)} \tilde{R}_i & \delta_i = 0, \end{cases}$$

and

$$\begin{aligned} \mu_i^{(t)} &= \mathbf{W}_i' \boldsymbol{\beta}^{(t)}, & \sigma_i^{(t)} &= \exp(\mathbf{Z}_i' \boldsymbol{\gamma}^{(t)}/2), & \tilde{Y}_i &= \frac{Y_i - \mu_i^{(t)}}{\sigma_i^{(t)}}, \\ f(a) &= \frac{\varphi(a)}{\Phi(-a)}, & g(a) &= 1 + \frac{a\varphi(a)}{\Phi(-a)}, \end{aligned}$$

and  $\varphi(a)$  and  $\Phi(a)$  are the PDF and CDF of a standard Normal distribution. Indeed, for  $\mathcal{Z} \sim \mathcal{N}(0, 1)$  we have  $f(a) = E[\mathcal{Z} \mid \mathcal{Z} > a]$  and  $g(a) = E[\mathcal{Z}^2 \mid \mathcal{Z} > a]$ .

- *M-step for  $\boldsymbol{\beta}$* : The conditional maximum  $\boldsymbol{\beta}^{(t+1)} = \arg \max_{\boldsymbol{\beta}} Q_t(\boldsymbol{\beta}, \boldsymbol{\gamma}^{(t)})$  is given by the weighted linear regression estimate:

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{W}' \boldsymbol{\Omega}_{(t)} \mathbf{W})^{-1} \mathbf{W}' \boldsymbol{\Omega}_{(t)} \tilde{\mathbf{R}}, \quad \text{where} \quad \boldsymbol{\Omega}_{(t)}^{-1} = \text{diag}((\sigma_1^{(t)})^2, \dots, (\sigma_n^{(t)})^2).$$

- *M-step for  $\boldsymbol{\gamma}$* : Similarly,  $\boldsymbol{\gamma}^{(t+1)}$  maximizes the objective function

$$Q_t(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\gamma}) = -\frac{1}{2} \left[ \sum_{i=1}^n \frac{\tilde{U}_i}{\exp(\mathbf{W}_i' \boldsymbol{\gamma})} + \mathbf{W}_i' \boldsymbol{\gamma} \right],$$

where  $\tilde{U}_i = \tilde{S}_i - 2\tilde{R}_i \mathbf{W}_i' \boldsymbol{\beta}^{(t+1)} + (\mathbf{W}_i' \boldsymbol{\beta}^{(t+1)})^2$ . Once again this corresponds to the likelihood of the Gamma GLM with log link function, which can be maximized using standard regression software.

## 2 Application to the Colon Cancer Study

The study of Laurie et al. (1989) and Moertel et al. (1990) is one of the first successful clinical trials of adjuvant chemotherapy for colon cancer. The dataset contained  $N = 888$  patients with colon carcinoma randomly assigned to the control group (no treatment) or one of two chemotherapy treatment groups: levamisol combined with fluorouracil or levamisole alone. In addition to the treatment group, 10 covariates (e.g., gender, age, severity of cancer) for each subject were also recorded. Over half the survival times in the sample were right-censored ( $N_{\text{cens}} = 458$ ).

As a basis of comparison for the HAFT model, we consider both the homoscedastic AFT with log-Normal survivals and a Cox proportional hazards (CPH) model. Stepwise regression based on the AIC was employed to select the covariates in the AFT and CPH models amongst all main effects and second order interactions. The HAFT model was given the same location covariates as its homoscedastic counterpart, and for simplicity we set the shape covariates to all the main effects. Parameter estimates for the fitted models are in Appendix A.

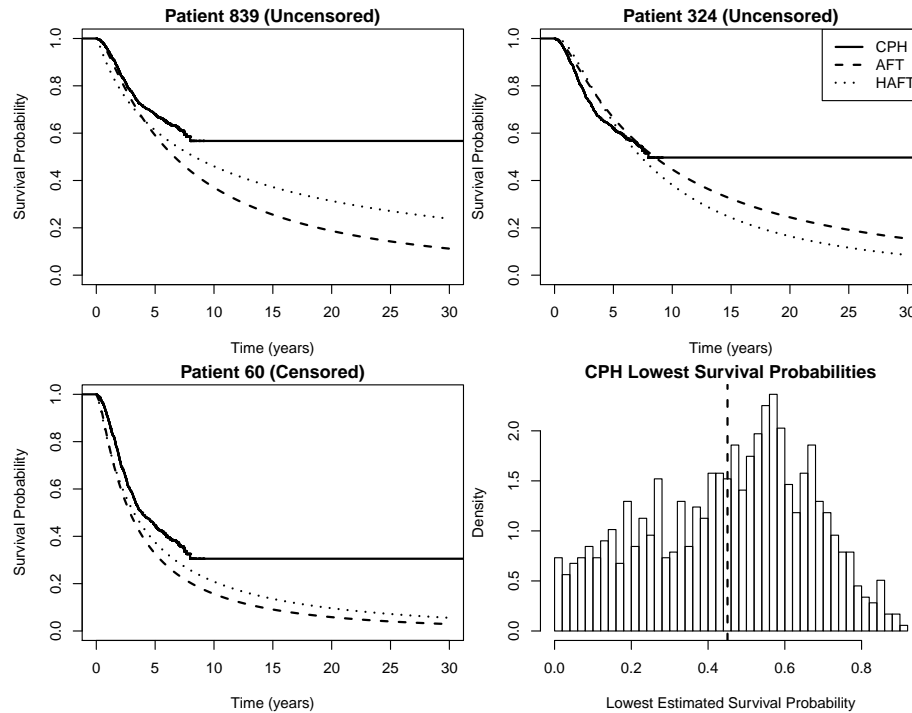


Figure 1: Left: Estimated survival curves for three randomly selected patients. Right: Lowest estimated survivals for the CPH model (mean indicated by dashed line).

## 2.1 Model Comparisons and Goodness-of-Fit

Figure 1 displays the estimated survival curves for all three models for several randomly selected subjects. Due to the high proportion of censored observations, the semi-parametric CPH model does not produce estimates for the upper survival quantiles. Indeed the CPH model truncates more than half of the predicted survival curves above 40% survival.

The AIC statistics for the parametric models are 7680.2 for AFT and 7671.0 for HAFT (the AIC for the CPH model is calculated from a partial likelihood and thus cannot be compared directly to the other two). To further compare AFT to its heteroscedastic extension, we consider the following goodness-of-fit tests for the model residuals.

For a given fitted model with parameters  $\boldsymbol{\theta}$ , we would like to compare the survival time  $T_i$  of each patient to its predictive distribution  $p(T_i | \mathbf{X}_i, \hat{\boldsymbol{\theta}})$ . In the absence of censoring, the HAFT model residuals are

$$\hat{\varepsilon}_i = \frac{R_i - \mathbf{W}_i' \hat{\boldsymbol{\beta}}}{\exp(\mathbf{Z}_i' \hat{\boldsymbol{\gamma}}/2)}.$$

With censoring, however, the observed data is not  $R_i$  but  $(Y_i, \delta_i)$ , with  $Y_i = \min(R_i, C_i)$  and  $\delta_i = \mathbf{1}[R_i < C_i]$ . A common approach to defining model residuals in the presence of censoring is to impute the missing survivals times (Hillis, 1995). That is, each censored observation is given a stochastic residual  $\tilde{\varepsilon}_i$ , computed as above but with  $\tilde{R}_i$  drawn from its truncated conditional distribution,

$$\tilde{R}_i \sim p(R | R > Y_i, \mathbf{X}_i, \hat{\boldsymbol{\theta}}).$$

The resulting Hillis residuals are approximately standard Normal under a correctly specified model. However, in the presence of heavy censoring as in our study, the Hillis residuals which are simulated from the posited model can easily overwhelm the uncensored data, and thus significantly decrease the power of goodness-of-fit tests.

Instead, we opted to fit a second parametric model to the conditional distribution of censoring times. While this requires additional assumptions, the large number of censored observations provided sufficient information to select AFT and HAFT candidate models for  $p(C | \mathbf{X})$ , exactly as for the survival distribution but with indicator  $1 - \delta$ .

Let  $f_{R|\mathbf{X}}(r|\mathbf{x})$ ,  $F_{R|\mathbf{X}}(r|\mathbf{x})$  and  $f_{C|\mathbf{X}}(c|\mathbf{x})$ ,  $F_{C|\mathbf{X}}(c|\mathbf{x})$  denote the condition PDF and CDF of survival and censoring distributions respectively. Then the conditional PDF of the observed time  $Y$  is

$$f_{Y|\delta, \mathbf{X}}(y|\delta = 1, \mathbf{X} = \mathbf{x}) \propto f_{R|\mathbf{X}}(y|\mathbf{x}) \cdot (1 - F_{C|\mathbf{X}}(y|\mathbf{x})), \quad (7)$$

$$f_{Y|\delta, \mathbf{X}}(y|\delta = 0, \mathbf{X} = \mathbf{x}) \propto f_{C|\mathbf{X}}(y|\mathbf{x}) \cdot (1 - F_{R|\mathbf{X}}(y|\mathbf{x})), \quad (8)$$

for uncensored and censored observations respectively.

While the conditional distributions for the AFT and HAFT models are Normal, distributions (7) and (8) are not. To construct residuals for this setting, we mapped each observation  $Y_i$  to its predicted Normal quantile:

$$\hat{\varepsilon}_i := \Phi^{-1}\left(P(Y \leq Y_i | \delta_i, \mathbf{X}_i, \hat{\boldsymbol{\theta}})\right), \quad (9)$$

where  $P(Y \leq y | \delta, \mathbf{X}, \boldsymbol{\theta})$  is the CDF associated with the PDFs in (7) and (8). The inner term in (9) thus corresponds to the probability integral transform of  $Y_i$ , such that the  $\hat{\varepsilon}_i$  are approximately standard Normal when both the survival and censoring models are correctly chosen.

## 2.2 Results

Figure 2 displays the observation times along with their predicted means and 95% prediction intervals. The data are grouped by censoring status, and the subjects are sorted on the  $x$ -axis in increasing order of the AFT model's predictions. (This is why the HAFT model predictions appear to be more erratic.)

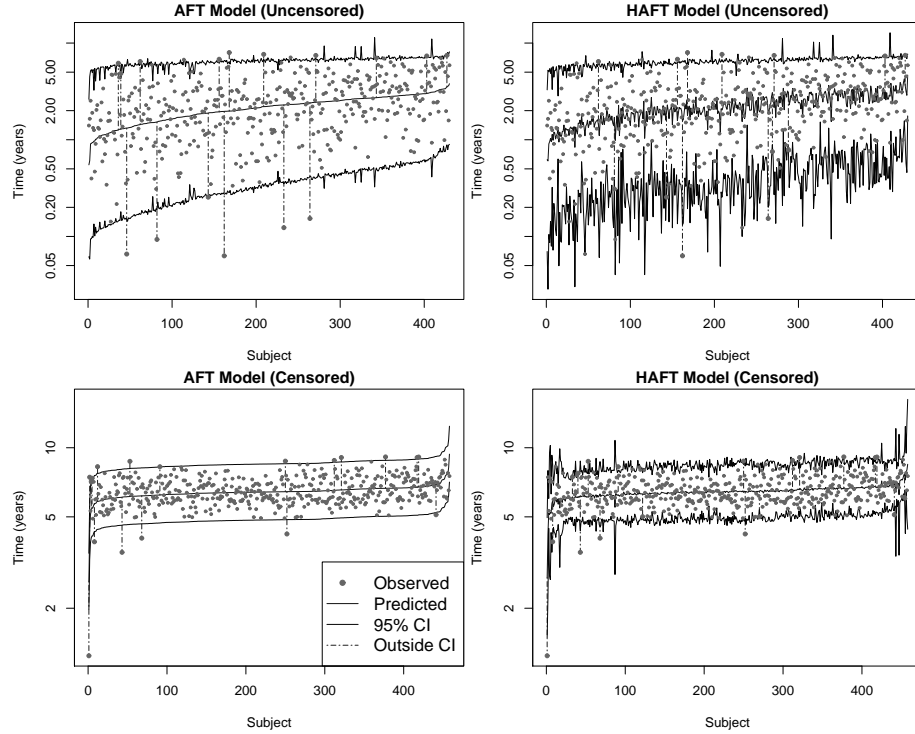


Figure 2: Observation times along with their predicted means and 95% prediction intervals.



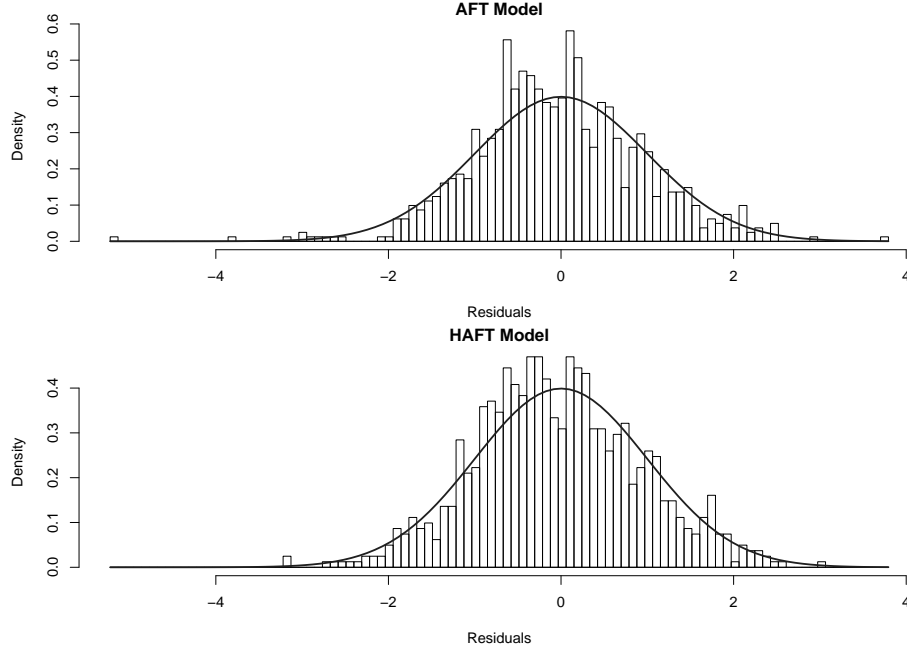


Figure 3: Residuals for AFT and HAFT models as computed by (9). The solid line corresponds to the  $\mathcal{N}(0, 1)$  residual distribution expected under the hypothesis that the model is correct.

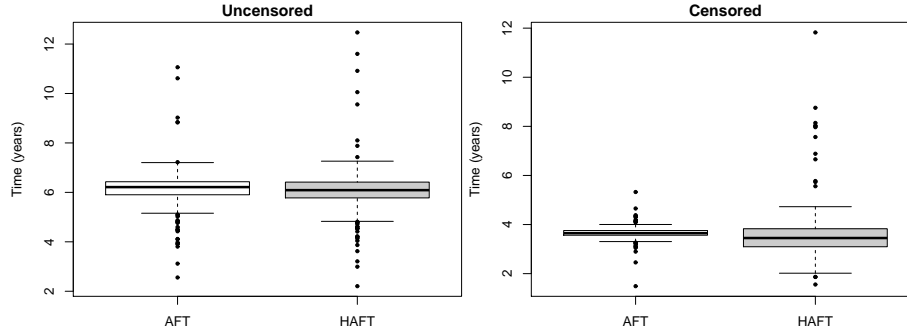


Figure 4: Widths of 95% prediction intervals for the AFT and HAFT models.

For the uncensored observations, the HAFT model has noticeably fewer outliers (indicated dashed lines). Both models have roughly the same outliers in the censored observations, but these are accompanied by wider prediction intervals with HAFT. This can also be seen from the model residuals  $\hat{\varepsilon}_i$ , calculated as in (9), which are shown in Figure 3. These suggest that accounting for het-

eroscedasticity in the AFT considerably improves the model’s fit, eliminating many of the extreme values.

While the HAFT model fits the data better by increasing its prediction intervals, it can also decrease prediction intervals for the non-extreme cases. The width of both models’ 95% prediction intervals are plotted in Figure 4. It is noteworthy that while the range of the HAFT prediction intervals is larger than AFT’s, its average prediction interval is actually smaller.

### 3 Discussion

The heteroscedastic AFT model we have proposed is a natural extension to its homoscedastic originator, and benefits from tractable computations in the presence of right-censoring. In an analysis of the colon cancer study which features heavy censoring, the HAFT model was found to substantially diminish the significance of model outliers, without increasing the average size of prediction intervals.

The results of this study are promising for the HAFT model, prompting several possible extensions to more complex models or with fewer assumptions. For instance, heavy-tailed residuals could be incorporated via the t-distribution, see Arellano-Valle et al. (2012). Alternately one might choose not to specify the residual distribution, in which case a number of semi-parametric homoscedastic AFT models (e.g., Buckley and James, 1979; Robins and Tsiatis, 1992; Zhang and Davidian, 2008; Zhou et al., 2012) can be adapted to the heteroscedastic setting. Similarly, it is possible to embed the HAFT model within more complex models to account for individual-level random effects or competing risks. It is hoped that the computational simplicity of the basic HAFT model can be leveraged to design effective Monte Carlo inference strategies in these more sophisticated settings.

**Acknowledgements** The work is supported by the Natural Sciences and Engineering Research Council of Canada grant number RGPIN-2014-04225.

## References

- Aitkin, M. (1981). A note on the regression analysis of censored data. *Technometrics*, 23(2):161–163.
- Angelis, D. D., Hall, P., and Young, G. (1993). Analytical and bootstrap approximations to estimator distributions in  $l_1$  regression. *Journal of the American Statistical Association*, 88(424):1310–1316.
- Arellano-Valle, R. B., Castro, L. M., González-Farías, G., and Muñoz-Gajardo, K. A. (2012). Student-t censored regression model: properties and inference. *Statistical Methods & Applications*, 21(4):453–473.
- Boscardin, W. J. and Gelman, A. (1996). Bayesian computation for parametric models of heteroscedasticity in the linear model. *Advances in Econometrics*, 11:87–110.
- Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.
- Cai, T. T. and Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, 36(5):2025–2054.
- Cox (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34(2):187–220.
- Hillis, S. L. (1995). Residual plots for the censored data linear regression model. *Statistics in medicine*, 14(18):2023–2036.
- Hougaard, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability*, 28(3):695–701.
- Hsieh, F. (1996). Empirical process approach in a two-sample location-scale model with censored data. *The Annals of Statistics*, 24(6):2705–2719.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley: Hoboken NJ.
- Keiding, N., Andersen, P. K., and Klein, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in medicine*, 16(2):215–224.
- Kocherginsky, M., He, X., and Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, 14(1):41–55.
- Koenker, R. (1994). Confidence intervals for regression quantiles. In *Asymptotic Statistics*, pages 349–359. Springer.
- Koenker, R. (2005). *Quantile Regression*. Cambridge Univ Press: Cambridge.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

- Laurie, J. A., Moertel, C., Fleming, T., Wieand, H., Leigh, J., Rubin, J., McCormack, G., Gerstner, J., Krook, J., and Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. the north central cancer treatment group and the mayo clinic. *Journal of Clinical Oncology*, 7(10):1447–1456.
- Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278.
- Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., Veeder, M. H., and Mailliard, J. A. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358.
- Moeschberger, M. and Klein, J. P. (1985). A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics*, 41(1):253–259.
- Müller, H. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics*, 15(2):610–625.
- Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, 74(2):221–232.
- Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime data analysis*, 7(1):55–64.
- Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.
- Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1):143–155.
- Robins, J. and Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2):311–319.
- Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society Series B*, 51(1):47–60.
- Smyth, G. K. (2002). An efficient algorithm for REML in heteroscedastic regression. *Journal of Computational and Graphical Statistics*, 11(4):836–847.
- Su, L., Zhao, Y., Yan, T., and Li, F. (2012). Local polynomial estimation of heteroscedasticity in a multivariate linear regression model and its applications in economics. *PloS one*, 7(9):e43719 1–13.
- Sy, J. P. and Taylor, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56(1):227–236.

- Verbyla, A. P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society Series B*, 55(2):493–508.
- Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11(14-15):1871–1879.
- Zeng, D. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society Series B*, 69(4):507–564.
- Zhang, J. and Peng, Y. (2009). Accelerated hazards mixture cure model. *Lifetime data analysis*, 15(4):455–467.
- Zhang, M. and Davidian, M. (2008). “Smooth” semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, 64(2):567–576.
- Zhou, M., Kim, M., and Bathke, A. (2012). Empirical likelihood analysis for the heteroscedastic accelerated failure time model. *Statistica Sinica*, 22:295–316.

## A Parameters of Fitted Survival Models

Table 1: Coefficients and Variances

	HAFt(sd) Location Parameters ( $\beta$ )	AFT(sd) Location Parameters ( $\beta$ )	CPH(sd) Hazard Parameters
(Intercept)	10.84(1.4e+00)	10.48(7.9e-01)	
rxLev	0.19(2.6e-02)	0.07(2.9e-02)	-0.24(1.9e-01)
rxLev+5FU	0.27(2.9e-02)	0.04(2.6e-02)	-0.19(1.8e-01)
sex	0.69(2.1e-01)	0.94(2.5e-01)	-1.04(5.2e-01)
age	-0.03(1.8e-04)	-0.02(1.5e-04)	0.03(1.4e-02)
obstruct	-0.25(2.2e-02)	-0.44(1.5e-02)	0.09(1.9e-01)
perfor	-0.30(3.7e-02)	-0.21(1.1e-01)	0.33(3.1e-01)
adhere	-1.21(5.2e-01)	-1.30(6.7e-01)	0.51(2.0e-01)
nodes	-0.17(1.4e-03)	-0.15(1.9e-03)	0.14(4.3e-02)
dif[moder]	-1.02(7.4e-01)	-0.74(6.6e-01)	1.24(9.4e-01)
dif[poor]	-2.81(9.5e-01)	-2.55(8.6e-01)	3.55(1.0e+00)
ext[muscle]	-0.64(8.3e-01)	-0.24(2.0e-01)	0.39(6.1e-01)
ext[serosa]	-1.01(8.1e-01)	-0.79(1.8e-01)	0.91(5.9e-01)
ext[cstruct]	-1.53(8.4e-01)	-1.25(2.3e-01)	1.28(6.2e-01)
surg	-0.20(1.2e-02)	-0.24(1.1e-02)	0.21(1.1e-01)
node4	-0.33(3.5e-02)	-0.44(3.6e-02)	0.48(1.9e-01)
I(nodes <sup>2</sup> )	0.004(1.5e-06)	0.004(3.0e-06)	-0.004(1.7e-03)
age:dif[moder]	0.02(1.8e-04)	0.02(1.7e-04)	-0.02(1.5e-02)
age:dif[poor]	0.04(2.4e-04)	0.04(2.2e-04)	-0.06(1.6e-02)
obstruct:perfor	0.73(1.5e-01)	1.19(3.5e-01)	-1.19(6.1e-01)
sex:age	-0.01(5.5e-05)	-0.02(6.4e-05)	0.02(8.2e-03)
rxLev:sex	-0.11(4.4e-02)	-0.13(5.3e-02)	0.10(2.3e-01)
rxLev:obstruct			0.61(2.8e-01)
rxLev+5FU:sex	0.44(6.1e-02)	0.39(5.6e-02)	-0.44(2.5e-01)
rxLev+5FU:obstruct			0.04(3.1e-01)
age:adhere	0.02(7.7e-05)	0.02(1.2e-04)	
adhere:dif[moder]	-0.05(2.3e-01)	-0.12(2.8e-01)	
adhere:dif[poor]	0.45(3.0e-01)	0.57(3.4e-01)	
adhere:nodes			-0.06(3.5e-02)
	Shape Parameters ( $\gamma$ )		
(Intercept)	0.01(1.1e+00)		
rxLev	0.29(3.2e-02)		
rxLev+5FU	0.66(3.8e-02)		
sex	0.02(2.4e-02)		
age	0.01(4.0e-05)		
obstruct	0.46(3.5e-02)		
perfor	-1.16(2.0e-01)		
adhere	-0.17(4.2e-02)		
nodes	-0.06(1.1e-03)		
dif[moder]	-0.25(7.3e-02)		
dif[poor]	0.09(9.5e-02)		
ext[muscle]	-0.35(1.1e+00)		
ext[serosa]	-0.00006(9.9e-01)		
ext[cstruct]	0.07(1.1e+00)		
surg	0.07(2.7e-02)		
node4	0.20(6.7e-02)		